

# Package ‘RIPSeekerData’

April 14, 2020

**Title** test data for RIPSeeker

**Version** 1.22.0

**Author** Yue Li

**Maintainer** Alejandro Reyes <reyes@embl.de>

**Description**

The RIP-seq data in BAM format are the test data for the package RIPSeeker. The data correspond to two RIP-seq experiments, namely PRC2 and CCNT1. Raw data from NCBI Gene Expression Omnibus under accession numbers GSE17064 for PRC2 and in-house for CCNT1.

**biocViews** ExperimentData, Genome, StemCell, SequencingData, RIPSeqData, GEO

**License** GPL-2

**Depends** RIPSeeker

**git\_url** <https://git.bioconductor.org/packages/RIPSeekerData>

**git\_branch** RELEASE\_3\_10

**git\_last\_commit** 5bb1758

**git\_last\_commit\_date** 2019-10-29

**Date/Publication** 2020-04-14

## R topics documented:

RIPSeekerData . . . . . 1

**Index** 3

---

RIPSeekerData                    *Alignment data in BAM files from RIP-seq experiments*

---

## Description

The RIP-seq data in BAM (binary version of the Sequence Alignment Map (SAM)) format are the test data for the package [RIPSeeker](#). The data correspond to two RIP-seq experiments, namely PRC2 and CCNT1 as detailed below. The BAM data will be imported, processed, and converted into [GappedAlignments](#) using RIPSeeker function [combineAlignGals](#).

## Usage

RIPSeekerData

## Format

BAM

## Details

PRC2 RIP-seq dataset: The RIP-seq data from Zhao et al. (2010) for Ezh2 (a PRC2 unique sub-unit) in mouse embryonic stem cell (mESC) were downloaded from Gene Expression Omnibus (GEO) (GSE17064). Briefly, there are in total five datasets. Two datasets correspond to the non-specific and specific negative controls using the antibody IgG and mutant mESC depleted of Ezh2 (Ezh2 *-/-*) (MT), respectively. Only the specific negative control was used in our test. The two and one remaining datasets correspond to the libraries constructed from two biological replicates of the wild type mESC. Notably, the library construction and *strand-specific* sequencing generated sequences from the opposite strand of the PRC2-bound RNA (Zhao et al., 2010), consequently, each read was treated as if it were reverse complemented. After the quality control (QC) and alignments, the technical replicates were merged, resulting in three test files - RIP-biorep1, RIP-biorep2, and CTL with 1,022,474, 442,030, and 208,445 reads mapped to unique loci of the mouse reference genome (mm9 build).

CCNT1 RIP-seq dataset: The data for CCNT1 were generated from two RIP-seq experiments. The pilot experiment generated 775,582 and 773,785 strand-specific raw reads, and 5,853 and 4,556 uniquely mapped read remain after the stringent QC for the CCNT1 and GFP control RIP RNA libraries, respectively. Same as in the PRC2 data, the reads came from the second strand of the cDNA synthesis opposite to the original RNA strand. The non-strand-specific library from the second screen has deeper coverage with 1,647,641 and 2,369,271 raw reads, and 26,859 and 45,024 uniquely aligned reads under QC for CCNT1 and GFP, respectively. Since the two experiments were performed with slightly different protocols, we treated them as two separate biological replicates for the following analyses.

For RIP-seq analysis using the [RIPSeeker](#) package, both of the PRC2 and CCNT1 BAM datasets will be imported, processed, and converted into [GappedAlignments](#) using RIPSeeker function [combineAlignGals](#).

## Source

Raw data from NCBI Gene Expression Omnibus under accession numbers GSE17064 for PRC2 and in-house for CCNT1.

## References

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., et al. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Molecular Cell*, 40(6), 939D953. doi:10.1016/j.molcel.2010.12.011

The RIPSeeker manuscript has been submitted to NAR for review.

# Index

## \*Topic **datasets**

RIPSeekerData, [1](#)

combineAlignGals, [1](#), [2](#)

GappedAlignments, [1](#), [2](#)

RIPSeeker, [1](#), [2](#)

RIPSeekerData, [1](#)